Stats 4CI3 - Assignment 3

Tommy Flynn (400121496)

February 15, 2021

Question 1:

We will compute the following integral: $\int_0^1 g(x)dx = \int_0^1 (\cos(30x) + \sin(10x))^2 \, dx$.

a) Using the integrate function we have

```
set.seed(24)
g = function(x) (cos(30*x) + sin(10*x))^2
integrate(g, 0 , 1)}
```

```
0.9867138 with absolute error < 2.3e-08
```

b) Using Monte Carlo integration we have

```
q = runif(n=10000, min=0, max=1)
1*mean(g(q))
```

```
[1] 0.9785487
```

Question 2:

a)

```
set.seed(24)
births = function(b,p){
  s = sample(c("Boy", "Girl"),size = b,
                 replace = T, prob=c(1-p,p))
  return(sum(s=="Girl"))
}
births(600,0.511)
```

```
[1] 313
```

By 1 simulation, we predict about 313 girl births.
b)

```
sim = replicate(1000, births(600,0.511))
mean(sim
hist(sim)
```

```
[1] 306.929
```

By 1000 simulations, we predict about 306.929 girl births. Below is the distribution of the results.

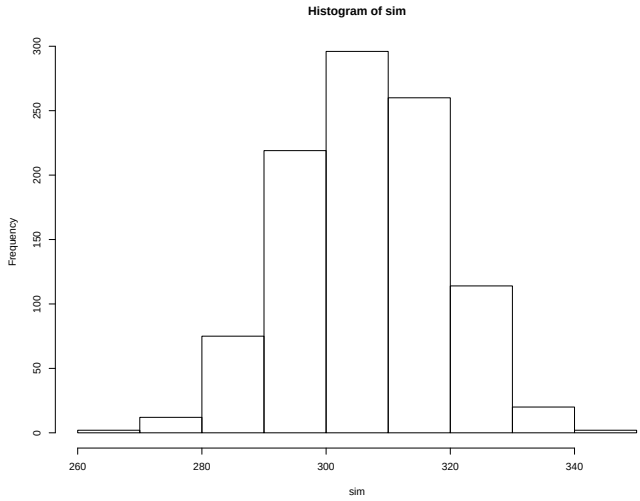

Figure 1: Histogram of girl births

Question 3:

We will simulate the linear model $y = 0.5 + 2x + \epsilon, \epsilon \sim N(0, 2^2), x \sim N(0, 1)$.
a) Simulation and scatter plot.

```
set.seed(24)
x = rnorm(1000,0,1)
e = rnorm(1000,0,2)
y = 0.5 + 2*x + e
plot(x,y)
```

b) The median.

```
abline(h=median(y),col="red")
median(y)
```

```
[1] 0.4172903
```

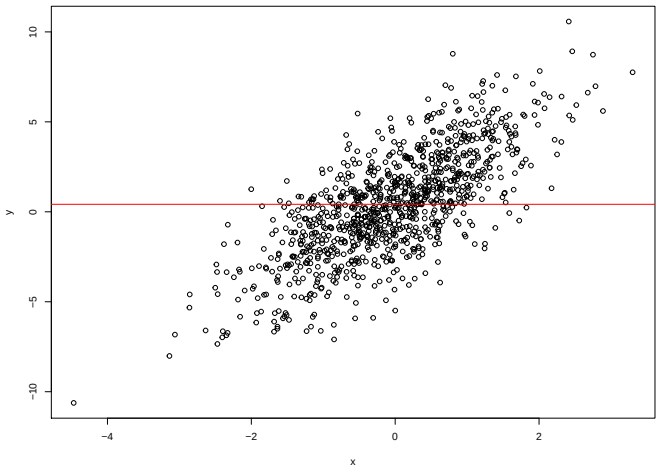

Figure 2: Scatter plot with median in red

Question 4:

First, here are functions that perform 1 simulation of A and B respectively.

```
dice_a = function(n){
  d_a = sample(1:6,size=n,replace=T)
  if (6 %in% d_a){
    return(1)}
  else {
    return(0)}
}

dice_b = function(n){
  d_b1 = sample(1:6,size=n,replace=T)
  d_b2 = sample(1:6,size=n,replace=T)
  d_b3 = d_b1 + d_b2
  if (12 %in% d_b3){
    return(1)}
  else {
    return(0)}
}
```

Now here is the repeated simulations and probabilities.

```
r = 100000
sim_a = replicate(r,dice_a(4))
p_a = mean(sim_a)
sim_b = replicate(r,dice_b(24))
p_b = mean(sim_b)
p_a
p_b
```

```
[1] 0.51829
[1] 0.49158
```

Therefore, by simulation, A is slightly more likely than B.

Question 5:

Here is the simulation and t-test.

```
set.seed(24)
x=rnorm(25,100,15)
test = t.test(x,mu=100)
```

a) The p-value is given by

```
test$p.value
```

```
[1] 0.3166587
```

b) The test statistic is given by

```
test$statistic
```

```
        t
-1.022678
```

c) We fail to reject the null hypothesis since the p-value= 0.3166587 is larger than $\alpha = 0.05$. Here is the simulation 10000 times.

```
pval = function(){
  x=rnorm(25,100,15)
  test = t.test(x,mu=100)
  return(test$p.value)
}
sim = replicate(10000,pval())
```

d) Here is the requested histogram.

```
hist(sim)
```

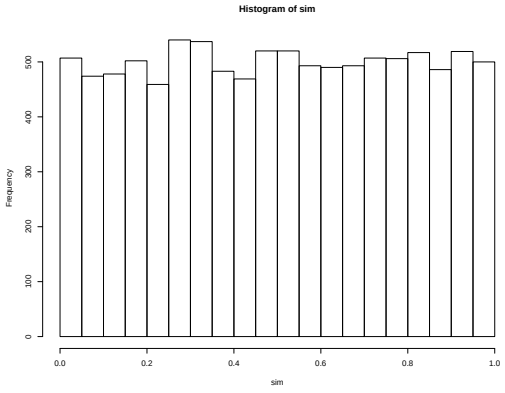

Figure 3: Histogram of p-values

e) Lastly, here is the proportion of p-values that are statistically significant.

```
length(which(sim<0.05))/10000
```

```
[1] 0.0507
```